

時間重要度をもつ位置依存コンテンツ検索のための P2P ネットワークに関する検討

A Study on P2P-Network to Search Location based Contents with Importance about Time

高橋 健太郎[◆] 中野 宏一[◆]
春本 要[◆] 西尾 章治郎[◇]

Kentaro TAKAHASHI Hirokazu NAKANO
Kaname HARUMOTO Shojiro NISHIO

店舗の営業時間やキャンペーン情報といった、位置と時間に依存する地域コンテンツを検索するサービスを P2P ネットワークで実現する上で、ピアの位置情報に基づいた論理ネットワークを構成すれば、検索対象の領域にクエリを効率的に配送できる。しかし、位置情報しか考慮されていないため、領域内の全てのピアにクエリが転送され、膨大なトラフィックが生じる。そこで本論文では、コンテンツの位置と時間帯別の重要度に着目したリンクの構築とルーティングの手法を提案する。提案手法では、位置依存の P2P ネットワークにおいて、 n ホップ以内のピアのもつコンテンツの位置と時間帯別の重要度をリンクテーブルに保持し、リンクを構築する。クエリとリンクテーブルからクエリの転送先を選択することで、クエリに対して高く適合するレスポンスの取得数を維持しつつトラフィックを削減する。

In this paper, a P2P (peer-to-peer) network is proposed that supports efficient query forwarding based on not only locations but also time-dependent importance of contents. Some location-based P2P networks are proposed so far, but they only consider locations. Therefore, a query is forwarded throughout the peers in the specified region, regardless of whether they have a relevant content or not. In our proposed method, location-dependent links are constructed using the existing method, and each peer also manages a link table that records neighboring peers' information that lie within n hops. The link table records time-dependent importance of content that the remote peers hold. By selectively forwarding a query based on the link table, query traffic is reduced while keeping the recall ratio of relevant contents.

◆ 学生会員 大阪大学大学院情報科学研究科マルチメディア工学専攻

{[takahashi.kentaro.nakano.hirokazu](mailto:takahashi.kentaro.nakano.hirokazu@ist.osaka-u.ac.jp)}@ist.osaka-u.ac.jp

◆ 正会員 大阪大学大学院工学研究科

harumoto@eng.osaka-u.ac.jp

◇ 正会員 大阪大学大学院情報科学研究科マルチメディア工学専攻 nishio@ist.osaka-u.ac.jp

1. はじめに

近年、携帯端末の普及やネットワーク技術の発展にともない、ユビキタス環境が整いつつある。ユビキタス環境ではインターネット上にこれまで以上に多様な情報（コンテンツ）が提供され、我々は日々の生活の中で様々な場所でそれらのコンテンツを活用できるようになると考えられる。特に、飲食や購買に際し、店舗コンテンツを推薦するようなサービスの需要が高まると考えられるが、その推薦の際に店舗の営業時間やキャンペーン情報といった時間を考慮して提示することにより、そのような推薦サービスを高度化できると考えられる。このとき、時間に依存したコンテンツの重要度を推薦に反映させる必要がある。

このようなサービスを実現することを考えた場合、コンテンツの管理面ではサーバ集中型のシステムが有効であるが、コンテンツ量の増大、クライアント数の増大、コンテンツのリアルタイム性を考慮した場合、情報発信源となる端末が直接情報を提供する P2P (Peer-to-Peer)型アーキテクチャを採用したシステムが有効である。

既存の P2P型アーキテクチャにおける代表的な検索手法であるフラッディングベースの検索には、取得できるコンテンツ数とトラフィックとの間にトレードオフの関係がある。この問題に対し、ピアの位置情報を考慮した論理ネットワークを形成する手法[1,2,3]が提案されており、これらを用いれば目的の領域へ効率的にクエリを転送することができる。しかし、これらの手法では位置情報以外の要素は考慮されていない。そのため、検索対象の領域内ではフラッディングを行うことになるため、クエリに適合しないコンテンツをもつピアにもクエリが転送されてしまい、トラフィックの問題は十分に解決されたとは言えない。

また、論理ネットワーク上で隣接するピアが保持するコンテンツの概要情報を用いてクエリの送信先を決定する手法[4]も提案されているが、扱うコンテンツをカテゴリに分類する必要があり、時間に依存したコンテンツの重要度を扱うには不適切である。

そこで本論文では、位置依存の P2P ネットワークを基盤とし、コンテンツのもつ時間帯別の重要度と位置情報に着目したクエリの転送方式を提案する。前提として、各ピアは自身のもつコンテンツの位置情報による論理ネットワークを構築しているものとする。その後、基盤とする P2P ネットワークのリンクに加え、 n ホップ内の他ピアとのリンクを構築するとともに、それらのピアがもつコンテンツの位置情報と時間帯別の重要度をリンクテーブルに保持する。ピアはクエリを受信すると、クエリの検索時間帯とリンクテーブルに基づいてクエリを送信するピアを選択し、クエリに対して適合度が低いピアへのクエリ転送を抑制することで、トラフィックを削減する。

2. 提案方式

2.1 想定環境

実際に地域コンテンツを検索するときには、位置情報以外に時間が重要となる。例えば、ユーザが「近くの飲食店」を探りたい場合、「近くの飲食店」のコンテンツ全てがユーザの意図する検索に適合するわけではない。ユーザが実際に訪れる時間に営業していない店舗のコンテンツは価値が低い。また、昼も夜も営業している店舗では、昼と夜で評判が違う

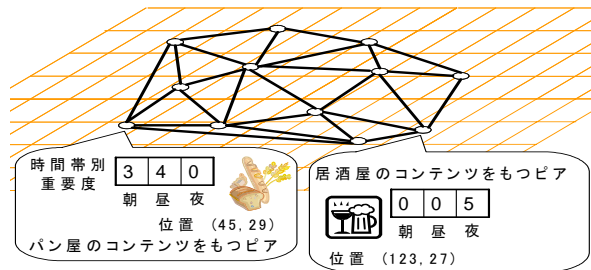


図1 コンテンツの位置情報による論理ネットワークの例
Fig.1 Example of a logical network based on content locations

こともある。このように、地域コンテンツの有用性は時間帯によって変化する。以降この有用性を重要度と定義する。検索するときには時間帯を指定することで、指定時間帯に重要度の高いコンテンツを検索することができる。

地域コンテンツは、店舗がキャンペーン広告のために作成したり、個人が時間を指定して不用品を引き取ってくれる相手を探すために作成したりする場合が考えられる。ここで、地域コンテンツには、位置情報と時間帯別の重要度が含まれているものとする。以降、この2つの情報をコンテンツの情報とする。また、各ピアはそれぞれただ1つのコンテンツを保持するものとする。

図1のように、ピアは位置情報によってドロネー図となる論理ネットワークを構築する。ただし、この位置情報とは実空間上においてピアが存在する地理的位置ではなく、コンテンツに付与された位置情報である。例えば、パン屋のコンテンツが店員の所持する携帯端末に存在するとき、店員が帰宅しても端末はパン屋の位置情報で論理ネットワークに参加する。ネットワークの分断はなく、全体で1つの論理ネットワークを形成しているものとする。このような論理ネットワークは、文献[1]や文献[2]で提案されている手法を利用するものとする。以下、この論理ネットワークを位置ネットワークと呼ぶ。

ピアがコンテンツを検索する際には、位置と時間を範囲指定したクエリを送信する。クエリには一意の識別子が付与されており、ピアは受信したクエリの識別子の履歴を作成することで、同じ識別子のクエリを重複して処理することはないものとする。また、クエリには経由したピアのアドレスが経路履歴として保持されており、経路履歴にあるピアにはクエリを送信しないことで、クエリのループを防止する。

2.2 リンクテーブル

位置ネットワークのトポロジであるドロネー図において、各ピアの平均リンク数は6である。この位置ネットワークにおいて単純にフラディングを行うと、クエリに適合しないピアや適合度が低いコンテンツを保持するピア（以下、このようなピアを適合度が低いピアと呼ぶ）にもクエリが転送されることになり、検索効率が低下する。ここで、隣接するピアが保持するコンテンツの情報を保持すれば、適合度が低いピアにクエリを転送しないことによってトラフィックを削減できると考えられるが、全ての隣接ピアの適合度が低い場合、クエリの転送を止めてしまうか、あるいはランダムに転送先を決定するしかない。ここで、位置ネットワーク上の隣接ピアを経由せずに2ホップ先あるいは3ホップ先のピアへ直接クエリを転送できれば、転送先の選択肢が広がり、トラフィックを削減できる可能性が高くなると考えられる。

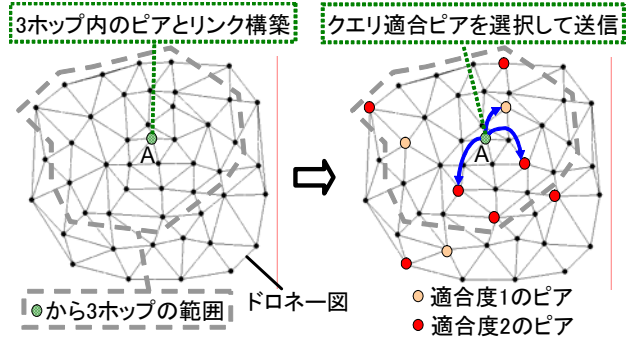


図2 リンクテーブルを用いたクエリの送信
Fig.2 Query forwarding using link table

そこで、ピアは位置ネットワーク上で n ホップ先までのピアともリンクを構築する。その際、ピアは他ピアの保持するコンテンツの位置情報および時間帯別重要度をリンクテーブルに保持する。図2は、位置ネットワークに対してピアAが3ホップ以内のピアをリンクテーブルに管理している例である。ピアAがクエリを送信するときには、リンクテーブルとクエリの指定時間帯から送信先を選択して、検索効率を向上させる。ただし、リンクテーブルで管理する最大ホップ数 n が大きくなるとテーブルサイズが増大するため、何らかの手段でテーブルサイズを制限しなければならない。テーブルサイズを現実的な大きさにすることと、トラフィック削減効果の向上とを併せて考慮し、3章の評価では $n = 3$ とした。このときの平均リンク数は40となる。

2.3 コンテンツのクエリに対する適合度

地域コンテンツの重要度は時間帯によって変化する。図1の例では、パン屋のコンテンツは昼に品揃えが充実するため、重要度は高くなり、夜は閉店するため、重要度は0となっている。このようにコンテンツの時間帯別重要度とクエリの指定時間帯から、コンテンツのクエリに対する価値として適合度を定義する。適合度は以下の式で計算する。

$$F = \sum I_t \quad (t \in \{c | Q_c = 1\})$$

ここで、 F は適合度、 I_t は時間帯 t でのコンテンツの重要度である。クエリが時間帯 c を指定時間帯にすると、 Q_c は1となる。適合度を計算するための t は、クエリの指定時間帯の集合である。つまり、適合度はクエリの指定時間帯でのコンテンツの時間帯別重要度の総和となる。図1の例ではパン屋のコンテンツの時間帯別重要度が朝に3、昼に4であるため、朝と昼を指定するクエリに対する適合度は7となる。

2.4 クエリ送信先の選択

検索領域外に存在するピアがクエリを受信した場合、リンクテーブル内のコンテンツの情報から最も検索領域に近くものを選び、そのコンテンツを保持するピアにクエリを転送する。クエリが検索領域内に到達すると複数リンクへのクエリ転送が開始される。以降、これをセレクトティブフラディングと呼ぶ。ただし、クエリは検索領域内から検索領域外へ送信されることはない。

同一ピアにクエリが複数送信されると、同一クエリを重複して受信してしまうため、不要なトラフィックが生じる。そこで、クエリの送信方向を制限することによって、同一クエリの重複受信を抑制する。各クエリには、セレクトティブフラディング開始ピアの位置情報が付加される。クエリを受

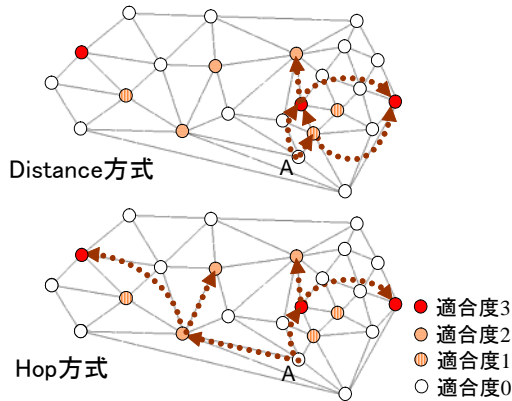


図3 提案方式におけるクエリの転送例
Fig.3 Query forwarding in the proposed methods

信したピアは、セレクトティブフラッディング開始地点に近くリンクにはクエリを送信しない。

クエリを送信するピアは、リンクテーブルとクエリの内容を参照して、リンク先のピアがもつコンテンツの適合度を計算する。しかし、適合度が0のピアにクエリを送信しないだけでは、トラフィックの削減効果が期待できない。送信クエリ数を限定し、適合度が高いピアの上位に送信することでトラフィックは削減できるが、適合度の高いピアにクエリが集中し、取得できるレスポンス数が低下する。そこで、適合度を基準として、より多くのレスポンスを取得できるように送信先を選択するためのスコアを定義する。クエリの転送先選択のためのスコア計算では、ピアのもつコンテンツの位置情報による距離を用いる計算式を Distance 方式、位置ネットワークにおけるピア間のホップ数を用いる計算式を Hop 方式として、2つの方式を比較する。

Distance 方式では、スコアを F/d で算出する。ここで、 F はクエリに対するコンテンツの適合度、 d はクエリ受信ピアと送信候補ピアの距離を示している。距離はピアのもつコンテンツの位置情報から決定される。スコア計算の結果、同じ適合度をもつリンクが複数ある場合は、距離が近いものが優先される。これは、リンク先ピアの適合度が全て同じ場合、距離が近いものにクエリを送信しなければ、送信方向の制限の影響で適合するコンテンツをもつピアにクエリが届かなくなるためである。図3は、ピアが右側には密に、左側には疎に位置する領域において、ピアAがクエリを送信する例を示している。位置的に近いピアのスコアが高くなりやすい Distance 方式では、分布密度の高い領域にクエリが転送されやすいため、検索領域内で、特定の領域にクエリが集中することがある。

Hop 方式では、スコアを F/h で算出する。ここで、 h はクエリ受信ピアから送信候補ピアまでのホップ数であり、1以上 n 以下となる。このようにクエリ受信ピアと送信候補ピア間の距離ではなくホップ数を用いることで、距離のスコアに対する影響を軽減でき、例えばクエリの検索領域内でのピアの分布に図3のような偏りがあっても、クエリが均等に広がると思われる。

クエリを送信するピアは、リンク先のピアそれぞれに対するスコアを算出した上で、スコアの上位 s 個のピアにクエリ

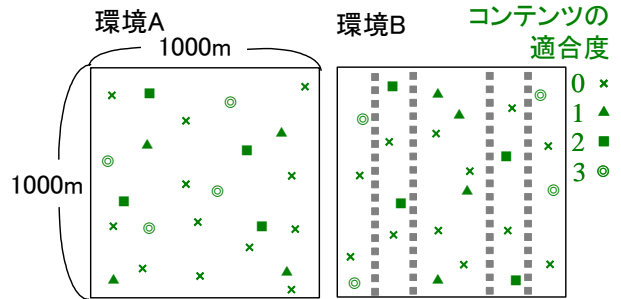


図4 領域内の適合度の分布
Fig.4 Distribution of relevance values

を送信する。このとき、適合度が0ではない送信可能ピアが2つしかなければ、 $s > 2$ でも2つのピアにしか送信しない。クエリに対するテーブル内のコンテンツの適合度が全て0であれば、クエリの転送が止まるのを防ぐため、テーブル内からクエリの経路履歴に含まれていないピアをランダムに1つ選択してクエリを送信する。テーブル内に送信可能なピアがなければ、クエリの転送は終了する。

3. 評価

提案手法をシミュレーションによって評価した。評価項目は、クエリあたりの平均ホップ数、および、取得したレスポンスの適合度ごとの取得率である。

3.1 評価環境

シミュレーション領域は1000平方メートルの領域とし、簡単のため、1日を10のタイムスロットに分割する。地域コンテンツの数の参考のために、Yahoo!エリア検索[5]において「北千里駅から半径0.5km」で検索したところ、検索結果は200件程度であった。想定するサービスでは、個人によるコンテンツの作成も考えられるため、ピア数は2倍の400と設定し、それぞれ1つのコンテンツをもつものとした。クエリは400個発生させ、検索領域におけるクエリの転送のみを評価するため、クエリの検索領域はエリア全体とした。

実際の地域コンテンツの分布では、時間帯別重要度の偏りが考えられる。例えば、繁華街に沿った店舗は居酒屋などが多く、夕方から深夜にかけて重要度が高くなるコンテンツが集まると考えられる。この場合、夜を指定するクエリに対する適合度の分布は、住宅街には低いものが集中し繁華街には高いものが集中する。そこで、検索領域内コンテンツのクエリに対する適合度の分布の偏りによる影響を評価するため、図4に示すような評価環境A、Bで提案手法を評価した。環境Aでは、発生するクエリに対して適合するコンテンツの分布は一様である。

環境Bでは図4のように、クエリに対する適合度に応じて、コンテンツの位置が異なる設定とした。エリア内で適合度が0のコンテンツは一様に分布する。また、縦に分割したエリアの中央に適合度1のコンテンツが分布し、適合度2,3のコンテンツは両側に分かれて分布する。両環境で、適合度別のコンテンツ数は同様であり、環境Aと比較することで適合度の分布の偏りによる影響を確認する。

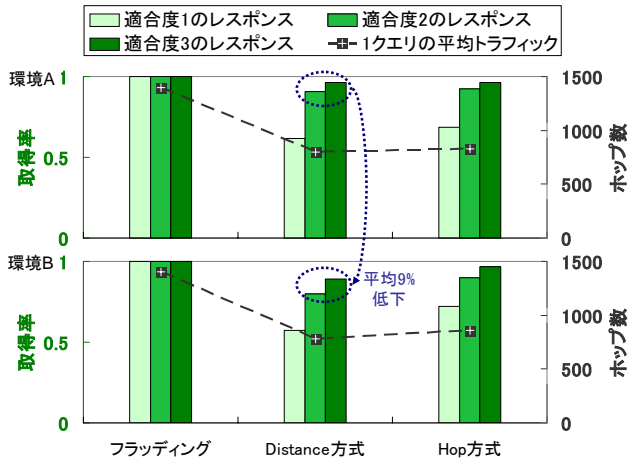


図5 トラフィックとレスポンスの適合度別取得率

シミュレーション環境において、ピアは位置情報による構造型のネットワークを構築しているため、通常のフラッディングでは同一ピアがクエリを複数受信することが多くなる。そこで、比較手法として用いるフラッディングは、提案手法と同様に送信方向の制限を設けた。

3.2 評価結果

環境 A, B における $s = 3$ としたときの実験結果のグラフを図 5 に示す。比較手法のフラッディングは、送信方向の制限によって、通常のフラッディングからトラフィックを 45% 削減したものである。制限を設けても、フラッディングではクエリに適合するコンテンツを全て取得している。

環境 A では、Distance 方式と Hop 方式の差はほとんど生じていない。しかし、フラッディングと比較するとトラフィックは 40% 以上減少しており、提案手法によるトラフィックの削減効果は明らかである。

環境 B において、Distance 方式の適合度上位 2 つのレスポンス取得率は、環境 A から平均 9% 低下している。これは、クエリが検索領域の端へ送られやすい環境 B では領域全体へクエリが配送されにくいためである。複数のピアへクエリを送信するとクエリの経路履歴が異なり、クエリを受信したピアが、既にそのクエリを受信済みのピアを転送先に選択する可能性がある。適合するピアが密集した領域では、位置的に近いピアのスコアが高くなると、密集した領域内でクエリを複数受信するピアが多くなり、領域外へクエリが広がりにくくなるのが分かる。Hop 方式では環境 A と比較して、トラフィックとリコール率に大きな変化が見られない。この結果から検索領域全体へクエリが広がりやすい Hop 方式の有効性が確認できる。高い適合度のピアが密集すると Distance 方式の性能が向上することも考えられ、今後検証する必要がある。

4. まとめと今後の課題

本論文では、コンテンツの位置と時間帯別の重要度に着目したリンクの構築とクエリのルーティング手法を提案した。また、シミュレーション実験による評価を行った。シミュレーション結果から、リンク先のピアがもつコンテンツの情報を保持することで、価値の高いレスポンスのリコール率を上げつつトラフィックの削減を行うことに成功した。

今後は、複数コンテンツをもつためのテーブルの集約と重

要度の連続性を考慮した手法への拡張を行う予定である。

【謝辞】

本研究は、総務省の委託研究「ユビキタスネットワーク認証・エージェント技術の研究開発」の成果の一部を活用したものである。また、本研究の一部は、文部科学省科学研究費補助金（課題番号 18700089）の助成によるものである。

【文献】

- [1] Araujo, F. and Rodrigues, L.: "GeoPeer: A location-aware peer-to-peer system", Proceedings of NCA '04, pp. 39-46 (2004).
- [2] 大西真晶, 源元佑太, 江口隆之, 加藤宏章, 西出 亮, 上島紳一: "ノード位置を用いた P2P モデルのためのドロネー図の自律分散生成アルゴリズム", 情報処理学会論文誌: データベース, Vol. 47, No. SIG4 (TOD29), pp. 51-64 (2006).
- [3] 金子 雄, 春本 要, 福村真哉, 下條真司, 西尾章治郎: "ユビキタス環境における端末の位置情報に基づく P2P ネットワーク", 情報処理学会論文誌: データベース, Vol. 46, No. SIG18 (TOD 28), pp. 1-15 (2005).
- [4] Crespo, A. and Garcia-Molina, H.: "Routing indices for peer-to-peer systems", Proceedings of ICDCS'02 (2002).
- [5] Yahoo Japan Corporation, "Yahoo! エリア検索", <URL: <http://area-search.yahoo.co.jp>>.

高橋 健太郎 Kentaro TAKAHASHI

大阪大学大学院情報科学研究科博士前期課程在学中。2007 年大阪大学工学部電子情報エネルギー工学科卒業。データベースシステム、マルチメディア情報システムなどの研究に従事。日本データベース学会学生会員。

中野 宏一 Hirokazu NAKANO

2007 年大阪大学大学院情報科学研究科博士前期課程修了。2005 年大阪大学工学部電子情報エネルギー工学科卒業。現在は東京都下水道局東部第一管理事務所勤務。

春本 要 Kaname HARUMOTO

大阪大学大学院工学研究科准教授。1994 年大阪大学大学院基礎工学研究科博士前期課程修了。博士 (工学)。データベースシステム、マルチメディア情報システムなどの研究に従事。電子情報通信学会、情報処理学会、IEEE 各会員。

西尾 章治郎 Shojiro NISHIO

大阪大学大学院情報科学研究科教授。1980 年京都大学大学院工学研究科博士後期課程了。工学博士。データベース、マルチメディアシステムの研究に従事。現在、Data & Knowledge Engineering 等の論文誌編集委員。本学会理事、電子情報通信学会、情報処理学会の各フェローを含め、ACM、IEEE など 8 学会の会員。